Real-Time Filtering of Vehicle Probe Data for Secondary Incident Prediction
Hyoshin Park*, Ali Haghani, Masoud Hamedi
Department of Civil and Environmental Engineering
University of Maryland, College Park, MD 20742
*Corresponding author: hspark@umd.edu

1 Introduction
Motor vehicle incidents remain a major cause of fatality, property loss, and disruption of service in transportation systems. They temporarily reduce roadway capacity to the existing traffic conditions from physical impedance in the travel lanes. The consequent speed reduction, queue formation, and rubbernecking foment additional incidents, referred to as secondary incidents. Contrary to the fact that each step of incident management system is documented and archived, secondary incidents at the incident location are not recorded. We need an assumption that defines and identifies the categories of secondary incidents. Better understanding and modeling of this phenomenon helps authorities to better prepare and respond to the event and mitigate congestion.

In recent years, the vehicle probe industry is emerging as a viable means to monitor network-wide traffic flow, delivering both speed and travel time information. This is a new opportunity to use real-time estimation of queue length to identify feasible area for secondary incidents. Such methodology has been absent in previous secondary incident studies.

The objective of this study is to develop real-time models for predicting secondary incidents using vehicle probe data. First, we estimate queue length to identify secondary incidents from archived incident data. Second, Bayesian Neural Network (BNN) models are used for the prediction of the secondary incidents. Finally, a pedagogical rule extraction technique is used to interpret BNN.

2 Feasibility Area
Previous identification of secondary incidents has focused on representing the temporal and spatial thresholds from the impact of primary events. Static thresholds were discussed by Karlaftis et al. (1999), Khattak et al. (2009), Hirunyanitiwattana and Mattingly (2006), Moore et al. (2004), Raud (1997), and Zhan et al. (2008). On the other hand, dynamic thresholds concluded that an incident should not classified as secondary when they occurred far from primary location of the event. Over the years, dynamics associated with traffic were considered by (i) simulation based approaches (Chou and Miller-Hooks, 2010 and Haghani et al., 2006); (ii) polynomial curves (Chilukuri and Sun, 2010); (iii) deterministic queuing methods (Khattak et al., 2012); (vi) three-phase traffic theory (Vlahogianni et al., 2012).
However, it is inappropriate to use a deterministic queuing approach for real-time application, since it assumes exact identification of arrival rate and capacity reduction (Fu and Rilett, 1997). Congestion caused by an accident might not classify the pronounced stop-and-go waves described as “wide moving jams” (Schönhof and Helbing, 2009). Since the above studies rely on loop detector data for the identification of secondary incidents, unsatisfactory raw data may decrease accuracy of the results. The novelty of this study rests in the attempt to use a probe vehicle technique to capture the dynamics of traffic evolution during primary events. Since no one has provided us with correct labels, it is assumed that individual component speeds may model some underlying set of hidden events with congested condition. A linear superposition of M component Gaussian Mixture model (GMM) is used to recognize speed and oscillation patterns of hidden events with congested condition, with D-dimensional continuous-valued data vector $x$, mixing coefficients $\omega_i$, mean vector $\mu$, and covariance matrix $\Sigma$.

$$P(x|\lambda) = \sum_{i=1}^{K} \frac{1}{2\pi^{D/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}.$$ 

Once the traffic jam is formed, rear-end crashes may occur at the tail of the queue as well as at the head of the queue due to large differences in speed (Marchesini and Weijermars, 2010). Some of cells recognized to be congestion-free by GMM have certain possibility to be in this group. To capture those phenomenon, we use an adjustment of the boxplot that considers medcouple (MC) to measure the skewness of a univariate sample from a continuous distribution $F$. Any cell with the speed value under: $Q_1$ (first quartile) – $1.5(MC)$ ($Q_3 - Q_1$) is regarded as congestion effected area. Queue lengths are estimated by adding up those congested cells. Figure 1 shows possible cases of false detection of secondary incidents ($S_3, S_4$) using method discussed by Hirunyaniwattana and Mattingly (2006). Khattak et al. (2012) cannot identify $S_3$ as secondary. Only proposed method can capture all correct secondary incidents including the opposite direction $S_2$. The incidents occurred in the years 2011 and 2012 at I-695 corridor are used in this study. Based on proposed framework, 112 (18%) are classified as primary-secondary pair among 614 crashes.

![Figure 1. Secondary incident classification by static and dynamic methods](image-url)
3 Prediction of Secondary Incidents
Previous crash analysis has difficulties with low sample mean and small sample size. Considering that incidents are random and rare, traditional neural networks may produce predicted values with unacceptable variances. A principled Bayesian learning approach to neural network is used in this study taking the notion that all sources of uncertainty in neural networks are expressed and measured by probabilities (Neal, 2011). A link function is explained by input $x_{ik}$ and parameters $(\alpha_p, \beta_j, \gamma_{jk})$, and used for predicting secondary incident value $\hat{y}_i$.

$$f_B(x_i, \theta) = \alpha_o + \sum_{k=1}^{p} (\alpha_k x_{ik}) + \sum_{j=1}^{m} \beta_j \tanh \left( \gamma_{jo} + \sum_{k=1}^{p} \gamma_{jk} x_{ik} \right),$$

$$\hat{y}_i = \int f_B(x_i, \theta) \times P(\theta | (x_1, y_1), \ldots, (x_n, y_n)) \ d\theta$$

A Hamilton Monte Carlo method is used for approximating the integral by samples drawn from the posterior distribution $P(\theta | (x_1, y_1), \ldots, (x_n, y_n))$, given secondary incident data $((x_1, y_1), \ldots, (x_n, y_n))$. A Hamilton equation, $H(q, p) = U(q) + p^T M^{-1} p / 2$, assumes that one can specify the exact position and momentum of a particle simultaneously at any point in time, where the potential energy is defined as $U(q)$ from the log probability density of the distribution for $q$, momentum vector $p$, and mass matrix $M$. The Metropolis hastings defines the Markov chain where the new sample $W(n + 1)$ is generated from the old sample $W(n)$. Twelve influential factors are used for modeling: incident duration, incident type, lane blockage, time of day, operational center, number of involved vehicles, heavy vehicles, traffic condition, response time, season, weather, and detection source. Note that predicted incident duration is used as one of the endogenous variables for prediction of secondary incidents. BNN yields satisfactory results based on mean absolute error, mean squared prediction error, and the proportion of underestimated values (0.27, 0.36, and 19.2%).

4 Characteristics of Secondary Incidents and Conclusion
If the reasoning behind the conclusion of learning system were more apparent, such that would lead to a better understanding of detected incidents, BNN model’s ability would be more worthwhile. However, previous neural networks have been regarded to be hard to interpret because learned knowledge were hidden from the user, and the models became just ‘black boxes’.

![Figure 2. Interpretation of Bayesian Neural Network for Secondary incidents](image)
A pedagogical rule extraction, proposed by Craven and Shavlik (1997), is applied to improve understanding of the secondary incidents by extracting comprehensible rules from the neural networks (Figure 2). Unlike most decision tree algorithms, we use a best-first expansion. Nodes with higher priorities are processed first with the greatest chance of increasing the information gain: \( G(n) = R(n)[1 - F(n)] \). \( R(n) \) is defined as the number of original samples reaching the node divided by total number of original training samples, whereas \( F(n) \) is the number of correctly classified samples divided by the number of all samples in the node. M-of-N splits are constructed by the heuristic search procedure which uses a beam-search method with a beam width of two. To avoid over-fitting, \( \chi^2 \) test is used to determine if the proposed change to the M-of-N test results in a significantly different partitioning of the instances than the partition induced by the test before the proposed change. Bonferroni correction is used to adjust the significant test downward for the individual tests.

Extracted decision trees provide a discovery and explanation of previously unknown relationships present in incident nature, and represent a series of decisions to assist emergency response personnel in better decision making. Traffic operators determine which incident cases have more priority under resource limitations.

References